

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: TAKAHASHI, Tohru
SERIZAWA, Nobufusa
KOISHI, Ryuta
KAWASHIMA, Ichiro
- (ii) TITLE OF INVENTION: EXPRESSION SYSTEMS UTILIZING
AUTOLYZING FUSION PROTEINS
AND A NOVEL REDUCING POLYPEPTIDE
- (iii) NUMBER OF SEQUENCES: 19
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Frishhauf, Holtz, Goodman, Langer & Chick, P.C.
(B) STREET: 767 Third Avenue-25th Floor
(C) CITY: New York
(D) STATE: New York
(E) COUNTRY: United States
(F) ZIP: 10017-2023
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.24
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/500,635
(B) FILING DATE: 11-JUL-1995
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 6-161053
(B) FILING DATE: 13-JUL-1994
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 6-218392
(B) FILING DATE: 13-SEP-1994
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 6-303809
(B) FILING DATE: 07-DEC-1994
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Goodman, Herbert
(B) REGISTRATION NUMBER: 17081
(C) REFERENCE/DOCKET NUMBER: 950376/HG
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212) 319-4900
(B) TELEFAX: (212) 319-5101
(C) TELEX: 236268

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1320 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

00842347 045501

TAT Tyr	CCT Pro	GAG Glu	AGA Arg	GAA Glu 165	GGT Gly	GAA Glu	CTC Leu	CGT Arg	CAA Gln 170	ACT Thr	GGA Gly	AAG Lys	GCA Ala	AGG Arg 175	TTA Leu	528
GTC Val	GAC Asp	CCA Pro	TCA Ser 180	GAG Glu	TTG Leu	CCC Pro	GCG Ala	CGG Arg 185	AAT Asn	GAG Glu	GAT Asp	ATT Ile 190	GAT Asp 190	GCA Ala	GAG Glu	576
TTT Phe	GAG Glu	AGT Ser 195	CTA Leu	AAT Asn	CGC Arg	ATA Ile 200	AGT Ser 200	GGT Gly	TTG Leu	CGC Arg	GAC Asp 205	TAT Tyr 205	AAT Asn	CCC Pro	ATT Ile	624
TCA Ser	CAA Gln 210	AAT Asn	GTT Val	TGC Cys	TTG Leu	CTA Leu 215	ACA Thr	AAT Asn	GAG Glu	TCA Ser	GAA Glu 220	GGC Gly	CAT His	AGA Arg	GAG Glu	672
AAG Lys 225	ATG Met	TTT Phe	GGA Gly	ATT Ile 230	GGA Gly 230	TAT Tyr	GGT Gly	TCA Ser	GTG Val	ATC Ile 235	ATT Ile 235	ACA Thr	AAT Asn	CAA Gln	CAT His 240	720
CTG Leu	TTC Phe	AGA Arg	AGG Arg	AAT Asn 245	AAT Asn	GGG Gly	GAG Glu	TTA Leu	TCA Ser 250	ATT Ile	CAA Gln	TCC Ser	AAG Lys	CAT His 255	GGC Gly	768
TAC Tyr	TTC Phe	AGA Arg	TGC Cys 260	CGC Arg	AAC Asn	ACC Thr	ACA Thr	AGC Ser 265	TTG Leu	AAG Lys	ATG Met	CTG Leu 270	CCT Pro	TTG Leu	GAG Glu	816
GGA Gly	CAT His	GAC Asp 275	ATT Ile	TTG Leu	TTG Leu	ATT Ile	CAG Gln 280	TTA Leu	CCA Pro	AGG Arg	GAC Asp 285	TTT Phe 285	CCA Pro	GTG Val	TTT Phe	864
CCA Pro	CAA Gln 290	AAG Lys	ATT Ile	CGC Arg	TTT Phe	AGG Arg 295	GAG Glu	CCA Pro	AGA Arg	GTG Val	GAT Asp 300	GAC Asp	AAA Lys	ATT Ile	GTT Val	912
TTG Leu 305	GTC Val	AGC Ser	ACA Thr	AAT Asn	TTC Phe 310	CAG Gln	GAA Glu	AAG Lys	AGT Ser	TCC Ser 315	TCG Ser	AGC Ser	ACG Thr	GTC Val	TCA Ser 320	960
GAG Glu	TCC Ser	AGT Ser	AAC Asn 325	ATT Ile	TCA Ser	AGA Arg	GTG Val	CAG Gln	TCA Ser 330	GCC Ala	AAT Asn	TTC Phe	TAC Tyr	AAG Lys 335	CAT His	1008
TGG Trp	ATC Ile	TCA Ser	ACA Thr 340	GTA Val	GCA Ala	GGA Gly	CAC His	TGT Cys 345	GGA Gly	AAC Asn	CCT Pro	ATG Met	GTT Val 350	TCG Ser	ACT Thr	1056
AAA Lys	GAT Asp	GGA Gly 355	TTT Phe	ATT Ile	GTA Val	GGT Gly 360	ATC Ile	CAC His	AGT Ser	CTT Leu	GCT Ala 365	TCA Ser 365	TTG Leu	ACA Thr	GGC Gly	1104
GAC Asp	GTT Val 370	AAC Asn	ATC Ile	TTC Phe	ACA Thr	AGC Ser 375	TTT Phe	CCG Pro	CCG Pro	CAG Gln	TTT Phe 380	GAG Glu	AAC Asn	AAA Lys	TAT Tyr	1152
CTA Leu 385	CAG Gln	AAG Lys	CTC Leu	AGT Ser	GAA Glu 390	CAC His	ACA Thr	TGG Trp	TGT Cys 395	AGT Ser 395	GGA Gly	TGG Trp	AAA Lys	CTA Leu	AAT Asn 400	1200
CTT Leu	GGA Gly	AAG Lys	ATT Ile 405	AGT Ser	TGG Trp	GGT Gly	GGA Gly	ATC Ile 410	AAC Asn 410	ATT Ile	GTG Val	GAG Glu	GAT Asp 415	GCA Ala 415	CCT Pro	1248

GAA GAG CCC TTT ATA ACA TCC AAG ATG GCA AGC CTT CTT AGT GAT TTG 1296
 Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu
 420 425 430

AAT TGT TCA TTC CAA GCA AGT GCG 1320
 Asn Cys Ser Phe Gln Ala Ser Ala
 435 440

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 440 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Clover Yellow Vein Virus

(ix) FEATURE:

- (A) NAME/KEY: mat_peptide
- (B) LOCATION: 4..437
- (D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys	Phe	Gln	Gly	Lys	Ser	Lys	Arg	Thr	Arg	Gln	Lys	Leu	Lys	Phe	Arg	1	5	10	15
Ala	Ala	Arg	Asp	Met	Lys	Asp	Arg	Tyr	Glu	Val	His	Ala	Asp	Glu	Gly	20	25	30	
Thr	Leu	Val	Glu	Asn	Phe	Gly	Thr	Arg	Tyr	Ser	Lys	Lys	Gly	Lys	Thr	35	40	45	
Lys	Gly	Thr	Val	Val	Gly	Leu	Gly	Ala	Lys	Thr	Arg	Arg	Phe	Thr	Asn	50	55	60	
Met	Tyr	Gly	Phe	Asp	Pro	Thr	Glu	Tyr	Ser	Phe	Ala	Arg	Tyr	Leu	Asp	65	70	75	80
Pro	Ile	Thr	Gly	Ala	Thr	Leu	Asp	Glu	Thr	Pro	Ile	His	Asn	Val	Asn	85	90	95	
Leu	Val	Ala	Glu	His	Phe	Gly	Asp	Ile	Arg	Leu	Asp	Met	Val	Asp	Lys	100	105	110	
Glu	Leu	Leu	Asp	Lys	Gln	His	Leu	Tyr	Leu	Lys	Arg	Pro	Ile	Glu	Cys	115	120	125	
Tyr	Phe	Val	Lys	Asp	Ala	Gly	Gln	Lys	Val	Met	Arg	Ile	Asp	Leu	Thr	130	135	140	
Pro	His	Asn	Pro	Leu	Leu	Ala	Ser	Asp	Val	Ser	Thr	Thr	Ile	Met	Gly	145	150	155	160
Tyr	Pro	Glu	Arg	Glu	Gly	Glu	Leu	Arg	Gln	Thr	Gly	Lys	Ala	Arg	Leu	165	170	175	
Val	Asp	Pro	Ser	Glu	Leu	Pro	Ala	Arg	Asn	Glu	Asp	Ile	Asp	Ala	Glu	180	185	190	

034434360

Phe Glu Ser Leu Asn Arg Ile Ser Gly Leu Arg Asp Tyr Asn Pro Ile
 195 200 205
 Ser Gln Asn Val Cys Leu Leu Thr Asn Glu Ser Glu Gly His Arg Glu
 210 215 220
 Lys Met Phe Gly Ile Gly Tyr Gly Ser Val Ile Ile Thr Asn Gln His
 225 230 235 240
 Leu Phe Arg Arg Asn Asn Gly Glu Leu Ser Ile Gln Ser Lys His Gly
 245 250 255
 Tyr Phe Arg Cys Arg Asn Thr Thr Ser Leu Lys Met Leu Pro Leu Glu
 260 265 270
 Gly His Asp Ile Leu Leu Ile Gln Leu Pro Arg Asp Phe Pro Val Phe
 275 280 285
 Pro Gln Lys Ile Arg Phe Arg Glu Pro Arg Val Asp Asp Lys Ile Val
 290 295 300
 Leu Val Ser Thr Asn Phe Gln Glu Lys Ser Ser Ser Ser Thr Val Ser
 305 310 315 320
 Glu Ser Ser Asn Ile Ser Arg Val Gln Ser Ala Asn Phe Tyr Lys His
 325 330 335
 Trp Ile Ser Thr Val Ala Gly His Cys Gly Asn Pro Met Val Ser Thr
 340 345 350
 Lys Asp Gly Phe Ile Val Gly Ile His Ser Leu Ala Ser Leu Thr Gly
 355 360 365
 Asp Val Asn Ile Phe Thr Ser Phe Pro Pro Gln Phe Glu Asn Lys Tyr
 370 375 380
 Leu Gln Lys Leu Ser Glu His Thr Trp Cys Ser Gly Trp Lys Leu Asn
 385 390 395 400
 Leu Gly Lys Ile Ser Trp Gly Gly Ile Asn Ile Val Glu Asp Ala Pro
 405 410 415
 Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu
 420 425 430
 Asn Cys Ser Phe Gln Ala Ser Ala
 435 440

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

THE

25

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

20

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

20

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

36

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCCAGGTGGT GGCCCAGGTG CTTGGAATGA ACAATT

36

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTGTCAGCAC ACCTGGGAGC TGTAGAGCTC

30

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ala Pro Gly Pro Pro Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids

[illegible]

(iii) HYPOTHETICAL: N

Pro Gly Pro Pro Pro Gly Pro
1 5

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1650 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: N

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (H) CELL LINE: KM-102

```
(ix) FEATURE:
      (A) NAME/KEY: CDS
      (B) LOCATION: 1..1647
      (D) OTHER INFORMATION:
```

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..69
(D) OTHER INFORMATION:

ATG TCA TGT GAG GAC GGT CGG GCC CTG GAA GGA ACG CTC TCG GAA TTG
Met Ser Cys Glu Asp Gly Arg Ala Leu Glu Gly Thr Leu Ser Glu Leu
-23 -20 -15 -10

GCC GCG GAA ACC GAT CTG CCC GTT GTG TTT GTG AAA CAG AGA AAG ATA
Ala Ala Glu Thr Asp Leu Pro Val Val Phe Val Lys Gln Arg Lys Ile
-5 1 5

GGC GGC CAT GGT CCA ACC TTG AAG GCT TAT CAG GAG GGC AGA CTT CAA
Gly Gly His Gly Pro Thr Leu Lys Ala Tyr Gln Glu Gly Arg Leu Gln
10 15 20 25

48

96

144

AAG Lys	CTA Leu	CTA Leu	AAA Lys	ATG Met 30	AAC Asn	GGC Gly	CCT Pro	GAA Glu	GAT Asp 35	CTT Leu	CCC Pro	AAG Lys	TCC Ser	TAT Tyr 40	GAC Asp	192
TAT Tyr	GAC Asp	CTT Leu	ATC Ile 45	ATC Ile	ATT Ile	GGA Gly	GGT Gly	GGC Gly 50	TCA Ser	GGA Gly	GGT Gly	CTG Leu	GCA Ala 55	GCT Ala	GCT Ala	240
AAG Lys	GAG Glu	GCA Ala 60	GCC Ala	CAA Gln	TAT Tyr	GGC Gly	AAG Lys 65	AAG Lys	GTG Val	ATG Met	GTC Val	CTG Leu 70	GAC Asp	TTT Phe	GTC Val	288
ACT Thr	CCC Pro 75	ACC Thr	CCT Pro	CTT Leu	GGA Gly	ACT Thr 80	AGA Arg	TGG Trp	GGT Gly	CTT Leu	GGA Gly 85	GGA Gly	ACA Thr	TGT Cys	GTG Val	336
AAT Asn 90	GTG Val	GGT Gly	TGC Cys	ATA Ile	CCT Pro 95	AAA Lys	AAA Lys	CTG Leu	ATG Met	CAT His 100	CAA Gln	GCA Ala	GCT Ala	TTG Leu 105	TTA Leu	384
GGA Gly	CAA Gln	GCC Ala	CTG Leu	CAA Gln 110	GAC Asp	TCT Ser	CGA Arg	AAT Asn	TAT Tyr 115	GGA Gly	TGG Trp	AAA Lys	GTC Val	GAG Glu 120	GAG Glu	432
ACA Thr	GTT Val	AAG Lys	CAT His 125	GAT Asp	TGG Trp	GAC Asp	AGA Arg	ATG Met 130	ATA Ile	GAA Glu	GCT Ala	GTA Val	CAG Gln 135	AAT Asn	CAC His	480
ATT Ile	GGC Gly	TCT Ser 140	TTG Leu	AAT Asn	TGG Trp	GGC Gly	TAC Tyr 145	CGA Arg	GTA Val	GCT Ala	CTG Leu	CGG Arg 150	GAG Glu	AAA Lys	AAA Lys	528
GTC Val 155	GTC Val	TAT Tyr	GAG Glu	AAT Asn	GCT Ala	TAT Tyr 160	GGG Gly	CAA Gln	TTT Phe	ATT Ile	GGT Gly 165	CCT Pro	CAC His	AGG Arg	ATT Ile	576
AAG Lys 170	GCA Ala	ACA Thr	AAT Asn	AAT Asn	AAA Lys 175	GGC Gly	AAA Lys	GAA Glu	AAA Lys	ATT Ile 180	TAT Tyr	TCA Ser	GCA Ala	GAG Glu 185	AGA Arg	624
TTT Phe	CTC Leu	ATT Ile	GCC Ala	ACT Thr 190	GGT Gly	GAA Glu	AGA Arg	CCA Pro	CGT Arg 195	TAC Tyr	TTG Leu	GGC Gly	ATC Ile	CCT Pro 200	GGT Gly	672
GAC Asp	AAA Lys	GAA Glu	TAC Tyr 205	TGC Cys	ATC Ile	AGC Ser	AGT Ser	GAT Asp 210	GAT Asp	CTT Leu	TTC Phe	TCC Ser	TTG Leu 215	CCT Pro	TAC Tyr	720
TGC Cys	CCG Pro	GGT Gly 220	AAG Lys	ACC Thr	CTG Leu	GTT Val	GTT Val 225	GGA Gly	GCA Ala	TCC Ser	TAT Tyr	GTC Val 230	GCT Ala	TTG Leu	GAG Glu	768
TGC Cys	GCT Ala 235	GGA Gly	TTT Phe	CTT Leu	GCT Ala	GGT Gly 240	ATT Ile	GGT Gly	TTA Leu	GAC Asp	GTC Val 245	ACT Thr	GTT Val	ATG Met	GTT Val	816
AGG Arg 250	TCC Ser	ATT Ile	CTT Leu	CTT Leu	AGA Arg 255	GGA Gly	TTT Phe	GAC Asp	CAG Gln	GAC Asp 260	ATG Met	GCC Ala	AAC Asn	AAA Lys	ATT Ile 265	864
GGT Gly	GAA Glu	CAC His	ATG Met	GAA Glu 270	GAA Glu	CAT His	GGC Gly	ATC Ile	AAG Lys 275	TTT Phe	ATA Ile	AGA Arg	CAG Gln	TTC Phe 280	GTA Val	912

CCA	ATT	AAA	GTT	GAA	CAA	ATT	GAA	GCA	GGG	ACA	CCA	GGC	CGA	CTC	AGA		960
Pro	Ile	Lys	Val 285	Glu	Gln	Ile	Glu	Ala	Gly	Thr	Pro	Gly	Arg	Leu	Arg		
								290					295				
GTA	GTA	GCT	CAG	TCC	ACC	AAT	AGT	GAG	GAA	ATC	ATT	GAA	GGA	GAA	TAT		1008
Val	Val	Ala	Gln	Ser	Thr	Asn	Ser 305	Glu	Glu	Ile	Ile	Glu	Gly	Glu	Tyr		
		300										310					
AAT	ACG	GTG	ATG	CTG	GCA	ATA	GGA	AGA	GAT	GCT	TGC	ACA	AGA	AAA	ATT		1056
Asn	Thr	Val	Met	Leu	Ala	Ile 320	Gly	Arg	Asp	Ala	Cys 325	Thr	Arg	Lys	Ile		
	315																
GGC	TTA	GAA	ACC	GTA	GGG	GTG	AAG	ATA	AAT	GAA	AAG	ACT	GGA	AAA	ATA		1104
Gly	Leu	Glu	Thr	Val	Gly 335	Val	Lys	Ile	Asn	Glu	Lys	Thr	Gly	Lys	Ile		
330										340					345		
CCT	GTC	ACA	GAT	GAA	GAA	CAG	ACC	AAT	GTG	CCT	TAC	ATC	TAT	GCC	ATT		1152
Pro	Val	Thr	Asp	Glu 350	Glu	Gln	Thr	Asn	Val 355	Pro	Tyr	Ile	Tyr	Ala 360	Ile		
GGC	GAT	ATA	TTG	GAG	GAT	AAG	GTG	GAG	CTC	ACC	CCA	GTT	GCA	ATC	CAG		1200
Gly	Asp	Ile	Leu 365	Glu	Asp	Lys	Val	Glu	Leu	Thr	Pro	Val	Ala	Ile	Gln		
								370					375				
GCA	GGA	AGA	TTG	CTG	GCT	CAG	AGG	CTC	TAT	GCA	GGT	TCC	ACT	GTC	AAG		1248
Ala	Gly	Arg	Leu	Leu	Ala	Gln 385	Arg	Leu	Tyr	Ala	Gly	Ser 390	Thr	Val	Lys		
		380															
TGT	GAC	TAT	GAA	AAT	GTT	CCA	ACC	ACT	GTA	TTT	ACT	CCT	TTG	GAA	TAT		1296
Cys	Asp	Tyr	Glu	Asn	Val	Pro 400	Thr	Thr	Val	Phe	Thr 405	Pro	Leu	Glu	Tyr		
	395																
GGT	GCT	TGT	GGC	CTT	TCT	GAG	GAG	AAA	GCT	GTG	GAG	AAG	TTT	GGG	GAA		1344
Gly	Ala	Cys	Gly	Leu	Ser 415	Glu	Glu	Lys	Ala	Val 420	Glu	Lys	Phe	Gly	Glu		
410														425			
GAA	AAT	ATT	GAG	GTT	TAC	CAT	AGT	TAC	TTT	TGG	CCA	TTG	GAA	TGG	ACG		1392
Glu	Asn	Ile	Glu	Val 430	Tyr	His	Ser	Tyr	Phe 435	Trp	Pro	Leu	Glu	Trp 440	Thr		
ATT	CCG	TCA	AGA	GAT	AAC	AAC	AAA	TGT	TAT	GCA	AAA	ATA	ATC	TGT	AAT		1440
Ile	Pro	Ser	Arg 445	Asp	Asn	Asn	Lys	Cys 450	Tyr	Ala	Lys	Ile	Ile 455	Cys	Asn		
ACT	AAA	GAC	AAT	GAA	CGT	GTT	GTG	GGC	TTT	CAC	GTA	CTG	GGT	CCA	AAT		1488
Thr	Lys	Asp 460	Asn	Glu	Arg	Val	Val 465	Gly	Phe	His	Val	Leu 470	Gly	Pro	Asn		
GCT	GGA	GAA	GTT	ACA	CAA	GGC	TTT	GCA	GCT	GCG	CTC	AAA	TGT	GGA	CTG		1536
Ala	Gly	Glu	Val	Thr	Gln 480	Gly	Phe	Ala	Ala	Ala	Leu 485	Lys	Cys	Gly			

[illegible]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 549 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met -23	Ser	Cys	Glu -20	Asp	Gly	Arg	Ala	Leu -15	Glu	Gly	Thr	Leu	Ser -10	Glu	Leu
Ala	Ala	Glu -5	Thr	Asp	Leu	Pro	Val 1	Val	Phe	Val	Lys 5	Gln	Arg	Lys	Ile
Gly 10	Gly	His	Gly	Pro	Thr 15	Leu	Lys	Ala	Tyr	Gln 20	Glu	Gly	Arg	Leu	Gln 25
Lys	Leu	Leu	Lys	Met 30	Asn	Gly	Pro	Glu	Asp 35	Leu	Pro	Lys	Ser	Tyr 40	Asp
Tyr	Asp	Leu	Ile 45	Ile	Ile	Gly	Gly	Gly 50	Ser	Gly	Gly	Leu	Ala 55	Ala	Ala
Lys	Glu	Ala 60	Ala	Gln	Tyr	Gly	Lys 65	Lys	Val	Met	Val	Leu 70	Asp	Phe	Val
Thr	Pro 75	Thr	Pro	Leu	Gly	Thr 80	Arg	Trp	Gly	Leu	Gly 85	Gly	Thr	Cys	Val
Asn 90	Val	Gly	Cys	Ile	Pro 95	Lys	Lys	Leu	Met	His 100	Gln	Ala	Ala	Leu	Leu 105
Gly	Gln	Ala	Leu	Gln 110	Asp	Ser	Arg	Asn	Tyr 115	Gly	Trp	Lys	Val	Glu 120	Glu
Thr	Val	Lys	His 125	Asp	Trp	Asp	Arg	Met 130	Ile	Glu	Ala	Val	Gln 135	Asn	His
Ile	Gly	Ser 140	Leu	Asn	Trp	Gly	Tyr 145	Arg	Val	Ala	Leu	Arg 150	Glu	Lys	Lys
Val 155	Val	Tyr	Glu	Asn	Ala	Tyr 160	Gly	Gln	Phe	Ile	Gly 165	Pro	His	Arg	Ile
Lys 170	Ala	Thr	Asn	Asn	Lys 175	Gly	Lys	Glu	Lys	Ile 180	Tyr	Ser	Ala	Glu	Arg 185
Phe	Leu	Ile	Ala	Thr 190	Gly	Glu	Arg	Pro	Arg 195	Tyr	Leu	Gly	Ile	Pro 200	Gly
Asp	Lys	Glu	Tyr 205	Cys	Ile	Ser	Ser	Asp 210	Asp	Leu	Phe	Ser	Leu 215	Pro	Tyr
Cys	Pro	Gly 220	Lys	Thr	Leu	Val	Val 225	Gly	Ala	Ser	Tyr	Val 230	Ala	Leu	Glu
Cys 235	Ala	Gly	Phe	Leu	Ala	Gly 240	Ile	Gly	Leu	Asp	Val 245	Thr	Val	Met	Val

Arg Ser Ile Leu Leu Arg Gly Phe Asp Gln Asp Met Ala Asn Lys Ile
 250 255 260 265
 Gly Glu His Met Glu Glu His Gly Ile Lys Phe Ile Arg Gln Phe Val
 270 275 280
 Pro Ile Lys Val Glu Gln Ile Glu Ala Gly Thr Pro Gly Arg Leu Arg
 285 290 295
 Val Val Ala Gln Ser Thr Asn Ser Glu Glu Ile Ile Glu Gly Glu Tyr
 300 305 310
 Asn Thr Val Met Leu Ala Ile Gly Arg Asp Ala Cys Thr Arg Lys Ile
 315 320 325
 Gly Leu Glu Thr Val Gly Val Lys Ile Asn Glu Lys Thr Gly Lys Ile
 330 335 340 345
 Pro Val Thr Asp Glu Glu Gln Thr Asn Val Pro Tyr Ile Tyr Ala Ile
 350 355 360
 Gly Asp Ile Leu Glu Asp Lys Val Glu Leu Thr Pro Val Ala Ile Gln
 365 370 375
 Ala Gly Arg Leu Leu Ala Gln Arg Leu Tyr Ala Gly Ser Thr Val Lys
 380 385 390
 Cys Asp Tyr Glu Asn Val Pro Thr Thr Val Phe Thr Pro Leu Glu Tyr
 395 400 405
 Gly Ala Cys Gly Leu Ser Glu Glu Lys Ala Val Glu Lys Phe Gly Glu
 410 415 420 425
 Glu Asn Ile Glu Val Tyr His Ser Tyr Phe Trp Pro Leu Glu Trp Thr
 430 435 440
 Ile Pro Ser Arg Asp Asn Asn Lys Cys Tyr Ala Lys Ile Ile Cys Asn
 445 450 455
 Thr Lys Asp Asn Glu Arg Val Val Gly Phe His Val Leu Gly Pro Asn
 460 465 470
 Ala Gly Glu Val Thr Gln Gly Phe Ala Ala Ala Leu Lys Cys Gly Leu
 475 480 485
 Thr Lys Lys Gln Leu Asp Ser Thr Ile Gly Ile His Pro Val Cys Ala
 490 495 500 505
 Glu Val Phe Thr Thr Leu Ser Val Thr Lys Arg Ser Gly Ala Ser Ile
 510 515 520
 Leu Gln Ala Gly Cys
 525

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TAAATAAATA AATAA

15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTAGCGCTCT GGGGCAAGCA TCCTCCAGGC TGGCTGCCAC CACCACCACC ACCACTGATC

60

TAGACT

66

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGTCAGCACA AATTTC

18

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AAACACAAC TGAATGAAC AATT

24

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCATTCCAAG TTGTGTTTGT GAAA

24

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CATAGGATGC TCCAACAA

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Asn Cys Ser Phe Gln Xaa

1

5